Differences in Differences (with a Time Series Detour)

Econ 140, Section 9
MM Chapter 5

Jonathan Old

Roadmap

1. IV revision
2. Time Series
3. Motivation for DiD
4. Rainfall IV: Checking relevance with DiD

IV revision

Group work

Group 1: We are interested in the effect of being in the army on crime. We instrument being in the army with a lottery (paper)

Group 2: We are interested in the effect of protestant religion on economic growth. We instrument protestantism in a region with the distance to Wittenberg (paper)

Group 3: We are interested in the effect of air pollution on mortality. We instrument local air pollution with wind direction (paper)

1. Relevance: Z must truly affect X
2. Independence/Exogeneity: Z is as good as randomly assigned
3. Exclusion restriction: The **only** way that Z affects Y is via X

Your job: Discuss whether these assumptions hold!

Any questions?

… Remember – Every question is useful!

Time Series

Why empirical microeconomists don't use time series

Spurious correlations here.

Unsolicited financial advice here.

A great idea to lose your tuition fees is here.

Time series topics

- Autoregressive and moving-average processes
- Serial correlation (of the errors)
- Unit roots
- Trends
- Detecting trend breaks
- ...

My prediction: Not the most important topic on the exam. But come to OH if you have questions or want to know more!

Motivation for DiD

Motivation for Differences in Differences

- In the first part of the course, we talked about comparison between groups or other units (cross-sectional data)
- But we have also seen some comparisons over time: time-series data
- Both of these have very big and obvious problems, but we can use them **together** using (panel data) in a powerful tool: Differences-in-differences!

How to get the causal effect of a treatment

In 2021, UC Berkeley offered free mental health coachings to students with pre-existing mental health issues. We want to evaluate the effect of this policy and we collect a depression score (0-10) for all students at UC Berkeley.

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention
- Estimated treatment effect? $6 - 5 = 1$

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention
- Estimated treatment effect? $6 - 5 = 1$
- Problems?

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention
- Estimated treatment effect? $6 - 5 = 1$
- Problems?
- **Identifying assumption**: The average difference between groups is due to the treatment only. Without the treatment, the average outcome of the treated group would have been equal to the average outcome of the control group.

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention
- Estimated treatment effect? $6 - 6 = 0$

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention
- Estimated treatment effect? $6 - 6 = 0$
- Problems?

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention
- Estimated treatment effect? $6 - 6 = 0$
- Problems?
- Identifying assumption: The average difference across time is due to the treatment only. Without the treatment, the average outcome of the treated group would not have changed.

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6 - 6) - (5 - 4) = (6 - 5) - (6 - 4) = -1$

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6 - 6) - (5 - 4) = (6 - 5) - (6 - 4) = -1$
- Identifying assumption: Parallel trends: Without the treatment, the average increase in the outcome of the treated would have been the same as the average increase in the outcome of the untreated.

Sidenote

... We can NOT observe the identifying assumption! We can find evidence for or against it, but we can never be sure!

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can calculate the DiD estimate in two ways (here: superscripts stand for treatment/control group, subscripts denote unit i and time $t \in \{0, 1\}$)

$$DiD = E[\ \underbrace{\left(Y_{i1}^{T} - Y_{i0}^{T}\right)}_{\text{Change for treated}} - \underbrace{\left(Y_{i1}^{C} - Y_{i0}^{C}\right)}_{\text{Change for untreated}}\ ]$$

$$= E[\ \underbrace{\left(Y_{i1}^{T} - Y_{i1}^{C}\right)}_{\text{After-difference}} - \underbrace{\left(Y_{i0}^{T} - Y_{i0}^{C}\right)}_{\text{Before-difference}}\ ]$$

DiD and Potential Outcomes

Let us look at the second one in more detail:

$$
\begin{aligned}
DiD &= E[\left(Y_{i1}^T - Y_{i1}^C\right) - \left(Y_{i0}^T - Y_{i0}^C\right)] \\
&= \underbrace{\left(E[Y_{i1}^T] - E[Y_{i1}^C]\right)}_{\text{After-difference}} - \underbrace{\left(E[Y_{i0}^T] - E[Y_{i0}^C]\right)}_{\text{Before-difference}}
\end{aligned}
$$

Compare this with out formula for selection bias (in time period 1):

$$
\begin{aligned}
\Delta &= E[Y_{i1}^T] - E[Y_{i1}^C] \\
&= \underbrace{E[\mathrm{Y}_{i1}^T(1) - \mathrm{Y}_{i1}^T(0)]}_{\text{ATT}} + \underbrace{\left(E[\mathrm{Y}_{i1}^T(0)] - E[\mathrm{Y}_{i1}^C(0)]\right)}_{\text{Selection Bias}}
\end{aligned}
$$

DiD and Potential Outcomes

Let us look at the second one in more detail:

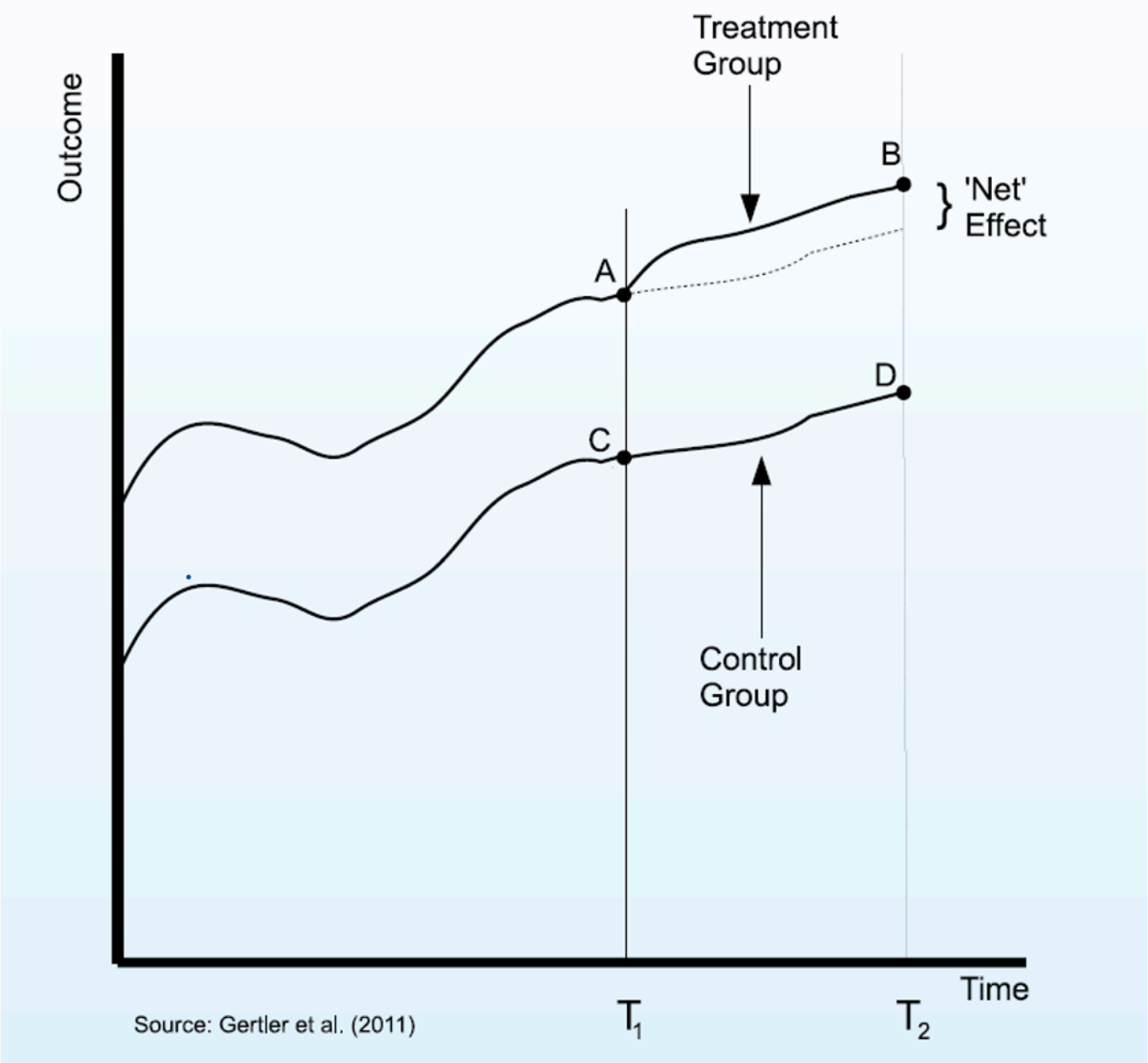
$$DiD = E[\left(Y_{i1}^T - Y_{i1}^C\right) - \left(Y_{i0}^T - Y_{i0}^C\right)]$$
$$= \underbrace{\left(E[Y_{i1}^T] - E[Y_{i1}^C]\right)}_{\text{After-difference}} - \underbrace{\left(E[Y_{i0}^T] - E[Y_{i0}^C]\right)}_{\text{Before-difference}}$$

Compare this with out formula for selection bias (at $t = 1$):

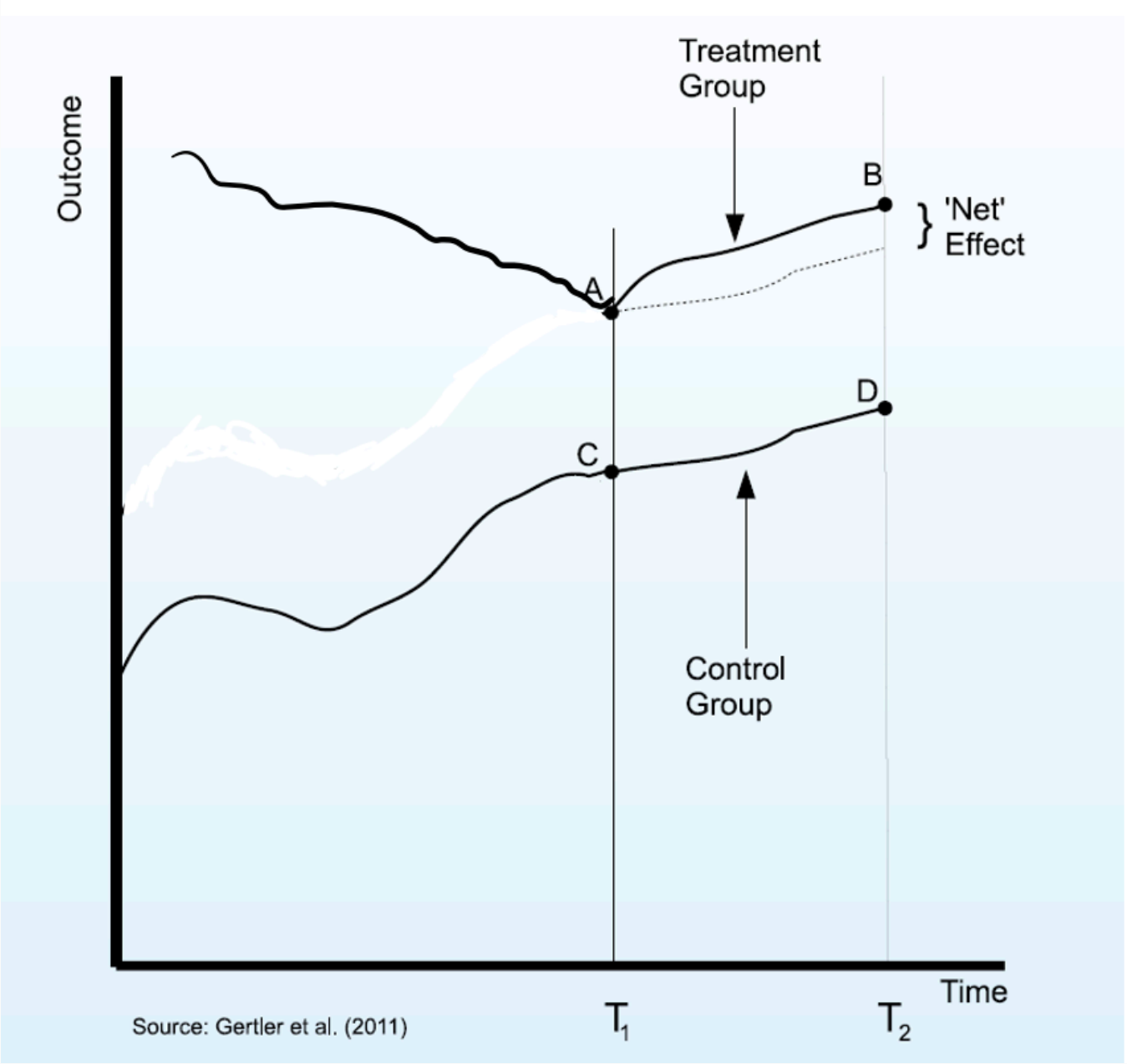
$$\Delta = E[Y_{i1}^T] - E[Y_{i1}^C]$$
$$= \underbrace{E[Y_{i1}^T(1) - Y_{i1}^T(0)]}_{\text{ATT}} + \underbrace{\left(E[Y_{i1}^T(0)] - E[Y_{i1}^C(0)]\right)}_{\text{Selection Bias}}$$

So DiD=ATT if the before-difference is equal to the selection bias in period 1: If the treatment group hadn't gotten the treatment (potential outcome = 0), the difference to the control group would have been the same as in period 0.

Parallel trends assumption



(Potential) violation of a parallel trends assumption



Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated		
No Free Mental Health: Untreated	α	

Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated		
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	$\alpha + \beta$	
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	$\alpha + \beta$	$\alpha + \beta + \gamma + \delta$
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Estimating DiD with regressions

We can set up a simple linear regression model to get the DiD estimate:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

With our data, we would get:

- $\alpha = 4$
- $\beta = 2$
- $\gamma = 1$
- $\delta = -1$

Any questions?

… Remember – Every question is useful!

Rainfall IV: Checking relevance with DiD

Rainfall and Income

You are hired by the Government of Ghana to study the impact of income on the level of education. Using data on rural villages, you estimate the following population regression using OLS:

$$Educ_i = \beta_0 + \beta_1 \text{ Income }_i + \beta_2 \text{ Pop }_i + \beta_3 \text{ School }_i + \beta_4 \text{ Age }_i + u_i$$

where Educ_i is average years of formal education in the village, Income_i is average annual income per capita in the village, Pop_i is the number of village residents, School_i is the number of schools in the village, and Age_i is the average age of the village population.

Explain what econometric problem is likely to arise that leads to biased and inconsistent estimates as a result of including Income as a regressor in the education regression as is done above.

IV regression

We want to know the effect of income on education and estimate

$$Educ_i = \beta_0 + \beta_1 \text{ Income }_i + \beta_2 \text{ Pop }_i + \beta_3 \text{ School }_i + \beta_4 \text{ Age }_i + u_i$$

You learn from Ghana's Minister of Agriculture that the country's citizens derive the bulk of their income from agriculture. As a result, you cleverly infer that average annual rainfall (Rainfall) may be a good instrument for income. You recall from your econometrics course that an instrument can be used in a procedure called Two Stage Least Squares that is designed to solve this econometric problem. Describe carefully the first of the two stages.

Checking Relevance with DiD

You want to check the Minister's suggestion that rainfall has an impact on incomes in Ghana. You have information on average annual incomes in 1996 and 1997 for two regions: the "coastal region," which had the same precipitation level in both years, and the "hill region," which experienced a 30% increase in rainfall. Comparing 1996 and 1997, income in the coastal region fell from 124 to 104, while income in the hill region fell from 98 to 96 . You also recall from your econometrics course that this situation might represent a "natural" or "quasi" experiment, allowing you to estimate the "treatment effect" of rainfall. Perform a difference in differences analysis of the effect of rainfall on average income. Summarize the analysis in a table.

DiD summary

Region	Rainfall	1996	1997
Coast (control)	No change	124	104
Hill (treatment)	+30%	98	96

This is equivalent to estimating:

$$I_{it} = \beta_0 + \beta_1 G_i + \beta_2 D_t + \beta_3 G_i \times D_t + u_i$$

where $G_i = 1$ if village is in Hills and $G_i = 0$ if village is on the Coast; $D_t = 1$ if year is 1997 and $D_t = 0$ if year is 1996. The differences in differences estimate of the rainfall effect is the OLS of β_3.